Constructive Asymptotic Equivalence of Density Estimation and Gaussian White Noise *

Michael Nussbaum and Jussi Klemelä
Weierstrass Institute, Berlin
Rolf Nevanlinna Institute, University of Helsinki

Abstract

A recipe is provided for producing, from a sequence of procedures in the Gaussian regression model, an asymptotically equivalent sequence in the density estimation model with i. i. d. observations. The recipe is, to put it roughly, to calculate square roots of normalised frequencies over certain intervals, add a small random distortion, and pretend these to be observations from a Gaussian discrete regression model.

Mathematics Subject Classifications: 62G07, 62B15, 62G20
Key Words: Nonparametric experiments, deficiency distance, Markov kernel, asymptotic minimax risk, curve estimation.

1 Introduction

In the first lecture notes of L. Le Cam from 1969 it is said: "En général une expérience est compliquée. Alors il faut l'approcher par une expérience plus simple." So the

*The research of Jussi Klemelä was financed by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" and to a smaller extent by Jenny and Antti Wihuri Foundation. The authors wish to thank G. Golubev for helpful discussions.

purpose is to approximate an experiment by a simpler one. The simplicity of an experiment is not defined precisely, but in a simple experiment one should be able to find optimal estimators, or at least asymptotically optimal estimators. By optimal estimator we mean either a minimax estimator, or also a Bayes estimator. Gaussian experiments are primary examples of simple experiments. Simplicity is also achieved by reducing the dimension of observations, as in the classical case of finding sufficient statistics whose dimension does not depend on the number of observations. Approximations are especially useful when one is able to transform the optimal procedures of a simple experiment to the optimal procedures of the more complicated experiment.

The precise sense of "approximate" is given by Le Cam's Δ-distance. When the Δ-distance between two experiments vanishes asymptotically, we say that the two experiments are asymptotically equivalent. The known results about asymptotic equivalence of experiments include the asymptotic equivalence of Gaussian discrete regression and Gaussian signal recovery by Brown and Low (1996), the asymptotic equivalence of density estimation with i. i. d. observations and Gaussian signal recovery by Nussbaum (1996), the asymptotic equivalence of non-Gaussian and Gaussian regression by Grama and Nussbaum (1996), and the asymptotic equivalence of spectral density and regression estimation by Golubev and Nussbaum (1998). With the exception of Brown and Low (1996) the previous articles do not give an explicit recipe for transforming a sequence of procedures of a simple experiment to the asymptotically equivalent sequence of the more complex experiment. In this article we study the density estimation with i. i. d. observations and give construction of Markov kernel which makes it possible to transform all the procedures constructed for the Gaussian experiment to the asymptotically equivalent procedures of the density experiment.

The basic idea might be described as "reduction to the Gaussian case by small distortions" (cp. Chapter 11.8 of Le Cam, 1986). Roughly speaking, the idea is as follows. Assume a simple model with real-valued parameter in which a real-valued statistic T_n is sufficient and asymptotically normal. Then, by small distortions, involving some additional randomization, one smoothes T_n in such a way that the "randomized" statistic has a density. This density then should reasonably converge to a normal density, entailing total variation convergence of the laws. The recipe is then "take the sufficient statistic, randomize, and pretend the resulting data are

Gaussian". Versions of such a theory (local and global) have been developed in parametric models by Le Cam (1986, Chapter 11.8) and by Müller (1981), and in a more indirect fashion in Shiryaev and Spokoiny (1993). In the framework of nonparametric asymptotic equivalence, a first result of this type was given by Brown and Low (1996b) for nongaussian regression. Our result pertains to the i.i.d. model, and takes the empirical distribution function as a sufficient statistic to start with. Apart form the initial smoothing, our proof is totally different from that of Brown and Low; it is inspired by Müller (1981), building upon rates of convergence in the functional central limit theorem. We do not attain the optimal smoothness index 1/2, but our method has a potential of being applicable wherever improved functional CLT's hold.

Before stating the main results, we have to give preliminary definitions.

Definition 1 **(i)** *A measurable space* $(\Omega, \mathcal{A})$ *is called* STANDARD BOREL *if there is a measurable space* $(\Omega_0, \mathcal{A}_0)$*, such that* $\Omega \in \mathcal{A}_0$*,* $\mathcal{A}$ *is the trace of* $\mathcal{A}_0$ *on* Ω*, i. e.* $\mathcal{A} = \{A \cap \Omega : A \in \mathcal{A}_0\}$*, and there is a metric on* Ω_0 *such that* Ω_0 *becomes a Polish space and* $\mathcal{A}_0$ *is the Borel sigma-algebra generated by the metric.*

(ii) *An experiment* $E = (\Omega, \mathcal{A}, (P_\theta, \theta \in \Theta))$ *is called* POLISH *if the measurable space* $(\Omega, \mathcal{A})$ *is standard Borel.*

(iii) *An experiment* $E = (\Omega, \mathcal{A}, (P_\theta, \theta \in \Theta))$ *is called* DOMINATED *if there exists a* σ*-finite measure* ν *on* $(\Omega, \mathcal{A})$ *such that* $P_\theta \ll \nu$*,* $\theta \in \Theta$*.*

(iv) *Given two measurable spaces* $(\Omega_i, \mathcal{A}_i)$*,* $i = 0, 1$*, a* MARKOV KERNEL *is a mapping* $K : \Omega_0 \times \mathcal{A}_1 \to [0, 1]$ *such that*

- **(a)** $K(\omega, \cdot)$ *is a probability measure on* $\mathcal{A}_1$ *for each* $\omega \in \Omega_0$*,*
- **(b)** $K(\cdot, A)$ *is a measurable function on* $(\Omega_0, \mathcal{A}_0)$ *for each* $A \in \mathcal{A}_1$*.*

(v) *The* TOTAL VARIANCE DISTANCE *between probability measures* P *and* Q *is*

$$\|P - Q\|_{TV} = \sup_{\|g\|_\infty \leq 1} \int g(dP - dQ).$$

Given experiments $E_i = (\Omega_i, \mathcal{A}_i, (P_{\theta,i}, \theta \in \Theta))$, $i = 0, 1$, the set of Markov kernels associated with the measurable spaces $(\Omega_i, \mathcal{A}_i)$ is denoted with $\mathcal{R}(E_0, E_1)$. Given $K \in$

$\mathcal{R}(E_0, E_1)$ and a probability measure P on $(\Omega_0, \mathcal{A}_0)$, KP is a probability measure on $(\Omega_1, \mathcal{A}_1)$, defined by $KP(A) = \int K(\omega, A)dP(\omega)$.

Definition 2 *Let $E_i = (\Omega_i, \mathcal{A}_i, (P_{\theta,i}, \theta \in \Theta))$, $i = 0, 1$, be two Polish dominated experiments.*

(i) *The* DEFICIENCY *of E_0 with respect to E_1 is*

$$\delta(E_0, E_1) = \inf_{K \in \mathcal{R}(E_0, E_1)} \sup_{\theta \in \Theta} \|KP_{\theta,0} - P_{\theta,1}\|_{TV}.$$

(ii) *The Δ-pseudodistance of E_0 and E_1 is*

$$\Delta(E_0, E_1) = \max\left\{\delta(E_0, E_1), \delta(E_1, E_0)\right\}.$$

We say that the sequence of experiments $E_{1,n}$ is asymptotically less informative than the sequence $E_{0,n}$, if $\delta(E_{0,n}, E_{1,n}) \to 0$. We say that the sequences of experiments $E_{0,n}$ and $E_{1,n}$ are asymptotically equivalent, if $\Delta(E_{0,n}, E_{1,n}) \to 0$. Now we are ready to give a definition of central importance to this article. According to this definition, a sequence of Markov kernels is asymptotically sufficient, if they achieve the infimum in the definition of the deficiency and the Markov kernels of the sequence are between two asymptotically equivalent experiments.

Definition 3 *Let $E_{i,n} = \left(\Omega_i^n, \mathcal{A}_i^n, (P_{\theta,i}^n, \theta \in \Theta)\right)$, $i = 0, 1$, be two Polish dominated experiments. A sequence of Markov kernels $K_n \in \mathcal{R}(E_{0,n}, E_{1,n})$ is* ASYMPTOTICALLY SUFFICIENT *if*

(i) $\lim_{n\to\infty} \sup_{\theta\in\Theta} \|K_n P_{\theta,0}^n - P_{\theta,1}^n\|_{TV} = 0$,

(ii) $\lim_{n\to\infty} \Delta(E_{0,n}, E_{1,n}) = 0$.

There are different ways of defining asymptotic sufficiency of a statistic or a sub-σ-field, stemming from Le Cam (1986). See for example Strasser (1985, Definition 81.3) or Laredo (1990, Corollary 1). One can define for example, that a statistic is asymptotically sufficient for an experiment if there is an experiment defined on the same probability space, which is asymptotically equivalent for the original experiment, and for which the statistic is exactly sufficient. This case is contained in Definition 3.

The usefulness of a sequence of asymptotically sufficient Markov kernels K_n lies in the fact that if we have a sequence of statistics $T_{n,1}$ for the experiments $E_{n,1}$, that is, measurable maps defined on $(\Omega_1^n, \mathcal{A}_1^n)$ to some other measurable space, then the sequence of statistics for the experiments $E_{n,0}$, defined by

$$T_{n,0}(\omega_0) = \int T_{n,1}(\omega_1) K_n(\omega_0, d\omega_1),$$

has the same asymptotic minimax risk and Bayes risk for bounded continuous loss functions, as the sequence $T_{n,1}$. Also, if we have a sequence of procedures $L_{n,1}$ for the experiments $E_{n,1}$, that is, Markov kernels defined on $(\Omega_1^n, \mathcal{A}_1^n)$ to some other measurable space, then the sequence of procedures for the experiments $E_{n,0}$, defined by

$$L_{n,0}(\omega_0, A) = \int L_{n,1}(\omega_1, A) K_n(\omega_0, d\omega_1),$$

has the same asymptotic minimax risk and Bayes risk for bounded continuous loss functions, as the sequence $L_{n,1}$.

Define a parameter space Σ of densities as follows. For $s \in]1,2]$ and $M > 0$, let $\Lambda^s(M)$ be the set of functions $[0,1] \to \mathbf{R}$ satisfying the condition

$$\left| f(t) - f(t_0) - f'(t_0)(t - t_0) \right| \leq M|t - t_0|^s. \tag{1}$$

For $\epsilon > 0$ define a set $\mathcal{F}_{\geq\epsilon}$ as the set of densities on $[0,1]$ bounded below by ϵ:

$$\mathcal{F}_{\geq\epsilon} = \left\{ f : \int_0^1 f = 1,\ f(x) \geq \epsilon,\ x \in [0,1] \right\}.$$

Define an a priori set, for given $s \in]1,2]$, $M > 0$, $\epsilon > 0$,

$$\Sigma = \Sigma_{s,M,\epsilon} = \Lambda^s(M) \cap \mathcal{F}_{\geq\epsilon}.$$

In Nussbaum (1996) it was shown that density estimation with i. i. d. observations and Gaussian signal recovery are asymptotically equivalent when the Hölder smoothness index $s > 1/2$ but we will have to assume that $s > 3/2$.

Let $X_1, \ldots, X_n$ be i. i. d. random variables with the density function $f \in \Sigma$. Let $P_{f,n}$ be the distribution of $X_1, \ldots, X_n$ and

$$E_{0,n} = \left([0,1]^n, \mathcal{B}^n_{[0,1]}, (P_{f,n}, f \in \Sigma) \right). \tag{2}$$

A simple experiment which is asymptotically equivalent to the experiment $E_{0,n}$ is the following experiment of discrete Gaussian regression. Let

$$y_i = \sqrt{f(s_i)} + \frac{1}{2}\left(\frac{N}{n}\right)^{1/2} \xi_i, \; i = 1, \ldots, N \tag{3}$$

where ξ_i are i. i. d. $N(0,1)$, $s_i = (t_{i-1} + t_i)/2$, and $t_i = i/N$. Let

$$Y_n = (y_i)_{i=1,\ldots,N}$$

and let $Q_{f,n}$ be the distribution of Y_n. Let $E_{1,n}$ be the corresponding experiment,

$$E_{1,n} = \left(\mathbf{R}^N, \mathcal{B}^N_{\mathbf{R}}, (Q_{f,n}, f \in \Sigma)\right). \tag{4}$$

The main theorem states roughly that the recipe for transforming density data to regression data without loosing information asymptotically, is to first calculate relative frequencies over the intervals $[t_{i-1}, t_i]$, multiply with N, add certain small randomization, and take square root from the result. Indeed, let

$$w_i = \int \psi_i d\hat{F}_n(1 + n^{-1/2} z_0) + n^{-1/2}\delta_n z_i, \;\; i = 1, \ldots, N \tag{5}$$

where

$$\psi_i = N I_{[t_{i-1}, t_i]},$$

$\hat{F}_n(t) = \sum_{i=1}^n I_{[0,t]}(X_i)$ is the empirical distribution function, z_i are i. i. d. $N(0,1)$, and $\delta_n > 0$.

Theorem 1 *Let $s > 3/2$ and $N = [n^{(1+\epsilon_1)/(2s)}]$ where $0 < \epsilon_1 < 2s/3 - 1$. The Markov kernel*

$$(X_1, \ldots, X_n) \longmapsto \left(\sqrt{\max\{w_i, 0\}}\right)_{i=1,\ldots,N} \tag{6}$$

from the density experiment $E_{0,n}$ to the Gaussian experiment $E_{1,n}$ is asymptotically sufficient, when we choose δ_n in the definition of w_i as

$$\delta_n = o(1), \;\; \delta_n^{-1} = o\left(n^{-3(1+\epsilon_1)/(4s)} n^{1/2} (\log n)^{-1}\right). \tag{7}$$

A recipe, almost similar to the one given in Theorem 1, was suggested by Donoho, Johnstone, Kerkyacharian, and Picard (1995, page 327).

Remark. Suppose we observe Y_n which has Poisson distribution with the intensity function nf, $f \in \Sigma$. Let

$$\chi_i = Y_n(i/n) - Y_n((i-1)/n), \;\; i = 1, \dots n,$$

and let $\tilde{F}_n(t)$ be the partial sum process formed from χ_i,

$$\tilde{F}_n(t) = n^{-1} \sum_{i=1}^{[nt]} \chi_i = n^{-1} Y_n([nt]/n).$$

Let $P_{f,n}$ be the distribution of $\tilde{F}_n$ and

$$\tilde{E}_{0,n} = (D[0,1], \mathcal{A}, (P_{f,n}, f \in \Sigma))$$

where $D[0,1]$ is the space of functions on $[0,1]$, continuous from the right and with the left side limits. Replacing the modified empirical distribution function $\hat{F}_n(1+n^{-1/2}z_0)$ by $\tilde{F}_n$ in the definition of w_i in (5), the statement of the Theorem 1 will hold also for the experiment $\tilde{E}_{0,n}$. $\square$

In the course of proving Theorem 1 we will also give an asymptotically sufficient Markov kernel from the density experiment to the following heteroscedastic Gaussian experiment. Let

$$\tilde{y}_i = f(s_i) + \left(\frac{N}{n}\right)^{1/2} f^{1/2}(s_i)\, \xi_i, \;\; i = 1, \dots, N$$

where ξ_i are i. i. d. $N(0,1)$, $s_i = (t_{i-1} + t_i)/2$, and $t_i = i/N$. Let

$$\tilde{Y}_n = (\tilde{y}_i)_{i=1,\dots,N}$$

and let $\tilde{Q}_{f,n}$ be the distribution of $\tilde{Y}_n$. Let $\tilde{E}_{1,n}$ be the corresponding experiment,

$$\tilde{E}_{1,n} = \left(\mathbf{R}^N, \mathcal{B}^N_{\mathbf{R}}, (\tilde{Q}_{f,n}, f \in \Sigma)\right). \tag{8}$$

The heteroscedastic experiment $\tilde{E}_{1,n}$ might be considered a bit more complicated as the homoscedastic experiment $E_{1,n}$. The following theorem states that the otherwise similar Markov kernel than defined in Theorem 1, but this time we do not take square roots, is asymptotically sufficient from the density experiment $E_{n,0}$ to the heteroscedastic Gaussian experiment $\tilde{E}_{n,1}$.

Theorem 2 *Let $s > 3/2$ and $N = [n^{(1+\epsilon_1)/(2s)}]$ where $0 < \epsilon_1 < 2s/3 - 1$. Let w_i be as defined in (5). The Markov kernel*

$$(X_1, \ldots, X_n) \longmapsto (w_i)_{i=1,\ldots,N} \tag{9}$$

from the density experiment $E_{0,n}$ to the heteroscedastic Gaussian experiment $\tilde{E}_{1,n}$ is asymptotically sufficient, when we choose δ_n in the definition of w_i as in (7).

1.1 Proof of Theorems 1 and 2

We will start by proving in Section 2 that if $\tilde{K}_n \in \mathcal{R}(E_{n,0}, \tilde{E}_{n,1})$ is the Markov kernel defined in (9), where $E_{n,0}$ is the density experiment defined in (2) and $\tilde{E}_{n,1}$ is the heteroscedastic Gaussian experiment defined in (8), then

$$\lim_{n\to\infty} \sup_{f\in\Sigma} \|\tilde{K}_n P_{f,n} - \tilde{Q}_{f,n}\|_{TV} = 0. \tag{10}$$

Secondly, we will prove in Section 3 that if $\bar{K}_n \in \mathcal{R}(\tilde{E}_{n,1}, E_{n,1})$ is the Markov kernel defined by

$$(\tilde{y}_i)_{i=1,\ldots,N} \longrightarrow \left(\sqrt{\max\{\tilde{y}_i, 0\}}\right)_{i=1,\ldots,N},$$

where $E_{n,1}$ is the homoscedastic Gaussian experiment defined in (4), then

$$\lim_{n\to\infty} \sup_{f\in\Sigma} \|\bar{K}_n \tilde{Q}_{f,n} - Q_{f,n}\|_{TV} = 0. \tag{11}$$

Thirdly, we will prove in Section 4 the following theorem which states that the homoscedastic Gaussian experiment $E_{1,n}$ is asymptotically equivalent to a continuous Gaussian white noise model.

Theorem 3 *Let*

$$E_{2,n} = \left(C[0,1], \mathcal{B}_{C[0,1]}, (Q^{(2)}_{f,n}, f \in \Sigma)\right)$$

where $Q^{(2)}_{f,n}$ is the distribution of the process

$$dX_n(t) = \sqrt{f(t)}dt + \frac{1}{2}n^{-1/2}dW(t), \ \ t \in [0,1]. \tag{12}$$

The experiments $E_{1,n}$ and $E_{2,n}$ are asymptotically equivalent, that is,

$$\Delta(E_{1,n}, E_{2,n}) \longrightarrow 0.$$

Theorem 1 is proved by noting that from (10) and (11) it follows that the condition (i) of Definition 3 is satisfied with the Markov kernel defined in (6). Thus we have also proved that the homoscedastic Gaussian experiment is asymptotically less informative than the density experiment, that is, $\delta(E_{0,n}, E_{1,n}) \to 0$. We need still to prove that $\delta(E_{1,n}, E_{0,n}) \to 0$. From Theorem 3 we have that $\delta(E_{1,n}, E_{2,n}) \to 0$. We know from Nussbaum (1996) that $\delta(E_{2,n}, E_{0,n}) \to 0$. Thus we have proved also $\delta(E_{1,n}, E_{0,n}) \to 0$, and thus $\Delta(E_{0,n}, E_{1,n}) \to 0$. Thus also the condition (ii) of Definition 3 is satisfied.

Theorem 2 is proved by noting that from (10) it follows that the condition (i) of Definition 3 is satisfied with the Markov kernel defined in (9). Thus we have also proved that the heteroscedastic Gaussian experiment is asymptotically less informative than the density experiment, that is, $\delta(E_{0,n}, \tilde{E}_{1,n}) \to 0$. We need still to prove that $\delta(\tilde{E}_{1,n}, E_{0,n}) \to 0$. From (11) it follows that $\delta(\tilde{E}_{1,n}, E_{1,n}) \to 0$. From Theorem 3 we have that $\delta(E_{1,n}, E_{2,n}) \to 0$. We know from Nussbaum (1996) that $\delta(E_{2,n}, E_{0,n}) \to 0$. Thus we have proved also $\delta(E_{1,n}, E_{0,n}) \to 0$, and thus $\Delta(E_{0,n}, E_{1,n}) \to 0$. Thus also the condition (ii) of Definition 3 is satisfied.

1.2 Inequalities for the Hellinger Distance

In the following, the Δ-distance will be mostly estimated using Hellinger distance. The Hellinger distance between probability measures P and Q is defined as

$$H^2(P,Q) = \int \left(f_P^{1/2} - f_Q^{1/2} \right)^2 .$$

where f_P and f_Q are probability densities of the distributions P and Q with respect to any dominating measure. Note that $\|P-Q\|_{TV} = \frac{1}{2}\int |f_P - f_Q|$. It holds that

$$\frac{1}{2} H^2(P,Q) \le \|P-Q\|_{TV} \le H(P,Q). \qquad (13)$$

When P and Q are product measures, $P = \prod_{i=1}^N p_i$, $Q = \prod_{i=1}^N q_i$, then it holds that

$$H^2(P,Q) \le 2 \sum_{i=1}^{N} H^2(p_i, q_i). \qquad (14)$$

The following inequalities hold:

$$H^2\left(N(\mu_1, \sigma^2), N(\mu_2, \sigma^2) \right) \le \frac{(\mu_1 - \mu_2)^2}{4\sigma^2}. \qquad (15)$$

From Golubev and Nussbaum (1998),

$$H^2\left(N(\mu,\Sigma_1),N(\mu,\Sigma_2)\right) \leq \frac{1}{16}\|\Sigma_1^{-1}\|\|\Sigma_2^{-1}\|\sum_{k,l}\left([\Sigma_1]_{k,l}-[\Sigma_2]_{k,l}\right)^2. \tag{16}$$

where

$$\|\Sigma\| = \sup_{\|x\|\leq 1} x'\Sigma x \leq Const \times \sum_{k,l}[\Sigma]^2_{kl}.$$

A special case of the previous inequality is

$$\begin{aligned} & H^2\left(N(\mu,\mathrm{diag}(\sigma^2_{1i})),N(\mu,\mathrm{diag}(\sigma^2_{2i}))\right) \\ \leq\ & Const \times \left(\sum_{i=1}^N \sigma_{1i}^{-2}\right)^{1/2}\left(\sum_{i=1}^N \sigma_{2i}^{-2}\right)^{1/2}\sum_{i=1}^N\left(\sigma^2_{1i}-\sigma^2_{2i}\right)^2. \end{aligned} \tag{17}$$

Finally we need that

$$\begin{aligned} H^2\left(\mathcal{L}\left(f(t)dt+n^{-1/2}dW(t)\right),\mathcal{L}\left(g(t)dt+n^{-1/2}dW(t)\right)\right) & \\ \leq Const \times n\int(f-g)^2. & \end{aligned} \tag{18}$$

2 The Heteroscedastic Experiment

In this section we will prove (10). That is, we will prove constructively that the heteroscedastic Gaussian experiment $\tilde{E}_{1,n}$ is asymptotically less informative than the density experiment $E_{0,n}$. The proof will take as a starting point the weak convergence

$$n^{1/2}(\hat{F}_n - F) \Rightarrow B\circ F, \tag{19}$$

where B is a Brownian bridge process, and $F(t) = \int_0^t f(s)ds$. From (19) one will move to the space of finite dimensional sequencies,

$$\int \psi_i d\hat{F}_n \approx \int \psi_i dF + n^{-1/2}\int \psi_i d(B\circ F),\ \ i=1,\ldots,N, \tag{20}$$

which suggests the accompanying heteroscedastic Gaussian experiment $\tilde{E}_{n,1}$, defined in (8).

In Section 2.1 we will prove the weak convergence with a rate for certain finite dimensional sequencies, in Section 2.2 we will strengthen the weak convergence to the convergence in the total variation norm, and in Section 2.3 we will move from (20) to the accompanying experiment $\tilde{E}_{n,1}$, defined in (8).

2.1 Weak Convergence

We will modify the empirical distribution function $\hat{F}_n$ to

$$\tilde{F}_n = \hat{F}_n(1 + n^{-1/2} z_0)$$

where $z_0 \sim N(0,1)$ is independent from the $X_1, \ldots, X_n$. Now we will have instead of (19),

$$n^{1/2}(\tilde{F}_n - F) \Rightarrow W \circ F \tag{21}$$

where W is a Wiener process, or Brownian motion, on $[0,1]$. We will start by proving a rate for the weak convergence of certain finite dimensional sequencies, to be defined in (25) and (26). For $x \in \mathbf{R}^N$, let the norm be

$$\|x\|_{seq} = \sup_{i=1,\ldots,N} N^{-1}|x_i|.$$

The rate of the weak convergence will be given in terms of the bounded Lipschitz distance. The bounded Lipschitz distance between probability distributions P and Q on $\mathbf{R}^N$ is defined by

$$\|P - Q\|_{BL} = \sup_{g \in BL} \int g(dP - dQ) \tag{22}$$

where

$$BL = \left\{ g : \mathbf{R}^N \longrightarrow \mathbf{R} \,\Big|\, \|g\|_\infty \le 1,\ \|g\|_L \le 1 \right\} \tag{23}$$

where

$$\|g\|_L = \sup_{x \neq y} \frac{|g(x) - g(y)|}{\|x - y\|_{seq}}.$$

The bounded Lipschitz distance metricises the weak convergence.

We can write

$$\begin{aligned} n^{1/2}(\tilde{F} - F) - W \circ F &= n^{1/2}(\hat{F} - F) - (W \circ F - z_0 F) + z_0(\hat{F} - F) \\ &= n^{1/2}(\hat{F} - F) - B \circ F + z_0(\hat{F} - F) \end{aligned}$$

where B is a Brownian bridge, which is defined in terms of the Wiener process W as $B(t) = W(t) - tW(1)$. By Bretagnolle and Massart (1989),

$$P\left(\|n^{1/2}(\hat{F} - F) - B \circ F\|_\infty > n^{-1/2}(x + 12 \log n) \right) \le 2 \exp(-x/6)$$

where $x > 0$ and for G in Skorohod space $D[0,1]$, $\|G\|_\infty = \sup_{t\in[0,1]}|G(t)|$. Thus, choosing for example $x = 6\log n$,

$$\begin{aligned}
&P\left(\|n^{1/2}(\tilde F - F) - W\circ F\|_\infty > 36n^{-1/2}\log n\right)\\
&\le \quad P\left(\|n^{1/2}(\hat F - F) - B\circ F\|_\infty + |z_0|\|\hat F - F\|_\infty > 36n^{-1/2}\log n\right)\\
&\le \quad P\left(\|n^{1/2}(\hat F - F) - B\circ F\|_\infty > 18n^{-1/2}\log n\right)\\
&\qquad +P\left(|z_0|\|\hat F - F\|_\infty > 18n^{-1/2}\log n\right)\\
&\le \quad Const\times n^{-r} \qquad (24)
\end{aligned}$$

where $r > 1/2$ and $Const$ does not depend on the unknown distribution F. Here we used an estimate

$$\begin{aligned}
&P\left(|z_0|\|\hat F - F\|_\infty > 18n^{-1/2}\log n\right)\\
&\le \quad P\left(|z_0|\|n^{1/2}(\hat F - F) - B\circ F\|_\infty > 9\log n\right) + P\left(|z_0|\|B\circ F\|_\infty > 9\log n\right)
\end{aligned}$$

and then the facts that

$$P\left(|z_0| > (9\log n)^{1/2}\right) \le n^{-9/2}$$

and

$$P\left(\|B\circ F\|_\infty > (9\log n)^{1/2}\right) \le n^{-9/2}$$

which follows from Talagrand (1988, Lemma 4). Now we will move to the space of sequences. Let

$$\Phi_n = \left(\int \psi_i d\left[n^{1/2}(\tilde F_n - F)\right]\right)_{i=1,\ldots,N} \qquad (25)$$

where, as before,

$$\psi_i(t) = NI_{[t_{i-1},t_i]}(t),$$

$t_i = i/N$. Let also

$$\Upsilon_n = \left(\int \psi_i d\,(W\circ F)\right)_{i=1,\ldots,N}. \qquad (26)$$

For $Y = n^{1/2}(\tilde F_n - F)$ and for $Y = W\circ F$ we have

$$\int \psi_i(t)dY(t) = N\int_{t_{i-1}}^{t_i} dY(t) = N(Y(t_{i-1}) - Y(t_i))$$

and thus

$$\left|\int \psi_i(t)dY(t)\right| \le 2N\|Y\|_\infty.$$

Thus

$$\|\Phi_n - \Upsilon_n\|_{seq} \leq 2\|n^{1/2}(\tilde{F}_n - F) - W \circ F\|_\infty$$

and thus, by (24),

$$\begin{aligned}
&P\left(\|\Phi_n - \Upsilon_n\|_{seq} > Const \times n^{-1/2}\log n\right) \\
&\leq \ P\left(\left\|\sqrt{n}(\tilde{F} - F) - W \circ F\right\|_\infty > Const \times n^{-1/2}\log n\right) \\
&\leq \ Const \times n^{-r}.
\end{aligned}$$

Now, for sufficiently large n,

$$\begin{aligned}
&\|\mathcal{L}(\Phi_n) - \mathcal{L}(\Upsilon_n)\|_{BL} \\
&= \ \sup_{g \in BL} E\left|g(\Phi_n) - g(\Upsilon_n)\right| \\
&\leq \ Const \times n^{-1/2}\log n + 2P\left(\|\Phi_n - \Upsilon_n\|_{seq} \geq Const \times n^{-1/2}\log n\right) \\
&\leq \ Const \times n^{-1/2}\log n + Const \times n^{-r} \\
&\leq \ Const \times n^{-1/2}\log n.
\end{aligned}$$

Because $Const$ does not depend on the density f, we have proved that

$$\sup_{f \in \Sigma} \|\mathcal{L}(\Phi_n) - \mathcal{L}(\Upsilon_n)\|_{BL} = O\left(n^{-1/2}\log n\right). \tag{27}$$

2.2 Total Variation Convergence

In this section we will strengthen the convergence in (27) to converegence in the total variation distance, but the convergence in the total variation distance will be without a rate.

Let Φ_n be as defined in (25) and let Υ_n be as defined in (26). We will add to Φ_n a certain small random disturbation U_n to strengthen the convergence in (27) to the convergence in the total variation norm. A similar kind of reasoning has been used by Müller (1981). Let

$$U_n = (\delta_n z_i)_{i=1,\ldots,N}$$

where z_i are i. i. d. $N(0,1)$ and δ_n is defined in (7). Now

$$\|\mathcal{L}(\Phi_n + U_n) - \mathcal{L}(\Upsilon_n)\|_{TV} \leq$$

$$\|\mathcal{L}(\Phi_n + U_n) - \mathcal{L}(\Upsilon_n + U_n)\|_{TV} \tag{28}$$

$$+\|\mathcal{L}(\Upsilon_n + U_n) - \mathcal{L}(\Upsilon_n)\|_{TV}. \tag{29}$$

The term (28) can be characterised as a variance term and (29) can be characterised as a bias term. To estimate (28),

$$\begin{aligned} &\|\mathcal{L}(\Phi_n + U_n) - \mathcal{L}(\Upsilon_n + U_n)\|_{TV} \\ &= \sup_{\|g\|_\infty \le 1} \left| E^{\Phi_n} E^{U_n} g(\Phi_n + U_n) - E^{\Upsilon_n} E^{U_n} g(\Upsilon_n + U_n) \right|. \end{aligned} \tag{30}$$

We should establish that when $g : \mathbf{R}^N \to \mathbf{R}$, $\|g\|_\infty \le 1$, then the function $x \mapsto E^{U_n} g(x + U_n)$ can be made a bounded Lipschitz $\mathbf{R}^N \to \mathbf{R}$ with $\|\cdot\|_{seq}$ norm.

Lemma 4 *Let $g : \mathbf{R}^N \to \mathbf{R}$ be such that $\|g\|_\infty \le 1$. Then for $x, y \in \mathbf{R}^N$,*

$$|E^{U_n} g(x + U_n) - E^{U_n} g(y + U_n)| \le L_n \|x - y\|_{seq}$$

where

$$L_n = 2^{-1/2} \delta_n^{-1} N^{3/2}.$$

Proof. Now, using (13), (14), and (15),

$$\begin{aligned} \left| E^{U_n} g(x + U_n) - E^{U_n} g(y + U_n) \right| &\le \sup_{\|g\|_\infty \le 1} \left| E^{U_n} g(x + U_n) - E^{U_n} g(y + U_n) \right| \\ &= \|\mathcal{L}(x + U_n) - \mathcal{L}(y + U_n)\|_{TV} \\ &\le H\left(\mathcal{L}(x + U_n), \mathcal{L}(y + U_n)\right) \\ &\le \left[2 \sum_{i=1}^N H^2\left(\mathcal{L}(x_i + \delta_n z_i), \mathcal{L}(y_i + \delta_n z_i)\right) \right]^{1/2} \\ &\le 2^{1/2} \left[\sum_{i=1}^N \frac{(x_i - y_i)^2}{4\delta_n^2} \right]^{1/2} \\ &\le 2^{-1/2} \delta_n^{-1} N^{3/2} \|x - y\|_{seq}. \end{aligned}$$

□

From the Lemma 4 we have that the function $x \mapsto E^{U_n} g(x + U_n)/\max\{L_n, 1\}$ is in BL, where BL was defined in (23). From (22), (27), and (30) we have an upper bound for (28):

$$\begin{aligned} \sup_{f \in \Sigma} \|\mathcal{L}(\Phi_n + U_n) - \mathcal{L}(\Upsilon_n + U_n)\|_{TV} &\le \max\{L_n, 1\} \sup_{f \in \Sigma} \|\mathcal{L}(\Phi_n) - \mathcal{L}(\Upsilon_n)\|_{BL} \\ &= \max\{L_n, 1\}\, O\left(n^{-1/2} \log n\right). \end{aligned} \tag{31}$$

Let us then estimate (29). We have

$$\begin{aligned}\|\mathcal{L}(\Upsilon_n + U_n) - \mathcal{L}(\Upsilon_n)\|_{TV} &= \frac{1}{2}\int |f_{\Upsilon_n} * f_{U_n} - f_{\Upsilon_n}| \\ &\le \frac{1}{4}\sum_{i=1}^{N}\int x_i^2 f_{U_n}(x)dx \int |D_i^2 f_{\Upsilon_n}|\end{aligned}$$

where the inequality is proved in Holmström and Klemelä (1992).

Now $\mathcal{L}(\Upsilon_n) = N(0, C)$ where for $i = 1, 2, \ldots, N$,

$$[C]_{ii} = Var\left(\int \psi_i d(W \circ F)\right) = \int \psi_i^2 dF = N \int \psi_i dF$$

and $[C]_{ij} = 0$ when $i \neq j$. Thus for $i = 1, \ldots, N$,

$$Const \times N \le [C]_{ii} \le Const \times N.$$

Thus

$$\left|[C^{-1}]_{ii}\right| \le Const \times N^{-1}$$

and $[C^{-1}]_{ij} = 0$ when $i \neq j$. Secondly,

$$\begin{aligned}D_i^2 f_{\Upsilon_n}(x) &= f_{\Upsilon_n}(x)\left[-[C^{-1}]_{ii} + \sum_{k,l} x_k x_l [C^{-1}]_{ik}[C^{-1}]_{il}\right] \\ &= f_{\Upsilon_n}(x)\left[-[C^{-1}]_{ii} + x_i^2 [C^{-1}]_{ii}^2\right]\end{aligned}$$

and thus

$$\int |D_i^2 f_{\Upsilon_n}| \le 2\left|[C^{-1}]_{ii}\right| \le Const \times N^{-1}$$

where $Const$ does not depend on f. Finally we get

$$\begin{aligned}\sup_{f\in\Sigma}\|\mathcal{L}(\Upsilon_n + U_n) - \mathcal{L}(\Upsilon_n)\|_{TV} &\le \frac{1}{4}\sup_{f\in\Sigma}\sum_{i=1}^{N}\int x_i^2 f_{U_n}(x)dx \int |D_i^2 f_{\Upsilon_n}| \\ &\le \sup_{f\in\Sigma}\frac{1}{2}\delta_n^2 \sum_{i=1}^{N}\left|[C^{-1}]_{ii}\right| \\ &\le Const \times \delta_n^2. \qquad (32)\end{aligned}$$

The two estimates, (31) for (28) and (32) for (29), will now be put together. For sufficiently large n:

$$\begin{aligned}\sup_{f\in\Sigma}\|\mathcal{L}(\Phi_n + U_n) - \mathcal{L}(\Upsilon_n)\|_{TV} &\le L_n O\left(n^{-1/2}\log n\right) + Const \times \delta_n^2 \\ &= Const \times \delta_n^{-1} N^{3/2} n^{-1/2}\log n + Const \times \delta_n^2.\end{aligned}$$

The upper bound converges to zero by the definitions of N and δ_n.

2.3 The Final Step

We have proved that

$$\sup_{f\in\Sigma} \|\mathcal{L}(\Phi_n + U_n) - \mathcal{L}(\Upsilon_n)\|_{TV} \longrightarrow 0$$

where Φ_n was defined in (25) and Υ_n was defined in (26). Let $L_{n,f} : \mathbf{R}^N \to \mathbf{R}^N$ be defined by

$$L_{n,f}(x) = \left(n^{-1/2}x_i + \int \psi_i dF\right)_{i=1,\ldots,N}.$$

Let

$$\Phi_n^{(1)} = L_{n,f}(\Phi_n) = \left(\int \psi_i d\tilde{F}_n\right)_{i=1,\ldots,N}$$

and

$$\Upsilon_n^{(1)} = L_{n,f}(\Upsilon_n) = \left(\int \psi_i dF + n^{-1/2}\int \psi_i d(W\circ F)\right)_{i=1,\ldots,N}.$$

Because the total variation distance is invariant under the affine-linear map $L_{n,f}$, it follows that

$$\sup_{f\in\Sigma} \|\mathcal{L}(\Phi_n^{(1)} + n^{-1/2}U_n) - \mathcal{L}(\Upsilon_n^{(1)})\|_{TV} \longrightarrow 0. \tag{33}$$

Let us denote

$$\Upsilon_n^{(2)} = \left(\int \psi_i dF + \left(\frac{N}{n}\right)^{1/2} f^{1/2}(s_i)\xi_i\right)_{i=1,\ldots,N}.$$

Applying (17) with $\sigma_{1i}^2 = n^{-1}\int \psi_i^2 dF = N\int \psi_i dF/n$ and $\sigma_{2i}^2 = Nf(s_i)/n$ we have that

$$\begin{aligned}
\|\mathcal{L}(\Upsilon_n^{(1)}) - \mathcal{L}(\Upsilon_n^{(2)})\|_{TV}^2 &\le H^2\left(\mathcal{L}(\Upsilon_n^{(1)}), \mathcal{L}(\Upsilon_n^{(2)})\right) \\
&\le Const \times \left(\sum_{i=1}^N \sigma_{1i}^{-2}\right)^{1/2} \left(\sum_{i=1}^N \sigma_{2i}^{-2}\right)^{1/2} \sum_{i=1}^N \left(\sigma_{1i}^2 - \sigma_{2i}^2\right)^2 \\
&\le Const \times n \sum_{i=1}^N \left(\frac{N}{n}\int \psi_i dF - \frac{N}{n} f(s_i)\right)^2 \\
&\le Const \times n^{-1}N^3 \sup_{i=1,\ldots,N} \left(\int \psi_i dF - f(s_i)\right)^2. \qquad (34)
\end{aligned}$$

According to (1) we can write $f(s_i+t) = f(s_i) + f'(s_i)t + R(t)$ where $|R(t)| \le M|t|^s$. Denote $K = I_{[-1/2,1/2]}$ so that $\psi_i(t) = NK(N(t-s_i))$. Then, because $\int K = 1$,

$\int tK(t)dt = 0$,

$$\begin{aligned}
\left|\int \psi_i f - f(s_i)\right| &= \left|N \int K(N(t-s_i))f(t)dt - f(s_i)\right| \\
&= \left|\int K(t)f(s_i + N^{-1}t)dt - f(s_i)\right| \\
&= \left|\int K(t)\left(f(s_i) + N^{-1}tf'(s_i) + R(N^{-1}t)\right)dt - f(s_i)\right| \\
&= M\int |N^{-1}t|^s |K(t)|dt \\
&\le Const \times N^{-s} \\
&= Const \times n^{-(1+\epsilon_1)/2}. \qquad (35)
\end{aligned}$$

Combining (34) and (35) we have,

$$\sup_{f\in\Sigma} \|\mathcal{L}(\Upsilon_n^{(1)}) - \mathcal{L}(\Upsilon_n^{(2)})\|_{TV} \longrightarrow 0. \qquad (36)$$

Let us denote, as before,

$$\tilde{Y}_n = \left(f(s_i) + \left(\frac{N}{n}\right)^{1/2} f^{1/2}(s_i)\xi_i\right)_{i=1,\dots,N}$$

where ξ_i are i. i. d. standard normal random variables. Now, using (13), (14), (15), and (35),

$$\begin{aligned}
\|\mathcal{L}(\Upsilon_n^{(2)}) - \mathcal{L}(\tilde{Y}_n)\|_{TV}^2 &\le H^2\left(\mathcal{L}(\Upsilon_n^{(2)}), \mathcal{L}(\tilde{Y}_n)\right) \\
&\le 2\sum_{i=1}^N H^2\left(\mathcal{L}((\Upsilon_n^{(2)})_i), \mathcal{L}((\tilde{Y}_n)_i)\right) \\
&\le 2\sum_{i=1}^N \frac{(\int \psi_i dF - f(s_i))^2}{4f(s_i)n^{-1}N} \\
&\le Const \times n \sup_{i=1,\dots,N}\left(\int \psi_i dF - f(s_i)\right)^2 \\
&\le Const \times nn^{-(1+\epsilon_1)} \longrightarrow 0. \qquad (37)
\end{aligned}$$

Combining the results (33), (36), and (37),

$$\sup_{f\in\Sigma} \|\mathcal{L}(\Phi_n^{(1)} + n^{-1/2}U_n) - \mathcal{L}(\tilde{Y}_n)\|_{TV} \longrightarrow 0.$$

The formula (10) is proved because

$$\Phi_n^{(1)} + n^{-1/2}U_n = (w_i)_{i=1,\dots,N} = \left(\int \psi_i d\tilde{F}_n + n^{-1/2}\delta_n z_i\right)_{i=1,\dots,N}.$$

3 Variance Stabilising Transform

Let us prove (11) in this section. That is, we will prove constructively that the homoscedastic Gaussian experiment $E_{1,n}$ is asymptotically less informative than the heteroscedastic Gaussian experiment $\tilde{E}_{1,n}$. The experiment generated by the observations

$$\sqrt{\max\{\tilde{y}_i, 0\}}, \ \ i = 1, \ldots, N$$

where

$$\tilde{y}_i = f(s_i) + \left(\frac{N}{n}\right)^{1/2} f^{1/2}(s_i)\xi_i$$

is equivalent to the experiment generated by the observations

$$z_i = \sqrt{y_i'}, \ \ i = 1, \ldots, N$$

where

$$y_i' = f(s_i) + \left(\frac{N}{n}\right)^{1/2} f^{1/2}(s_i)\xi_i',$$

$$\xi_i' = \begin{cases} \xi_i, & \text{when } \xi_i \geq -h \\ 0, & \text{otherwise} \end{cases}$$

where $\xi_i \sim N(0,1)$ are i. i. d. and

$$h = h(i,n,f) = \left(\frac{n}{N}\right)^{1/2} f^{1/2}(s_i).$$

Let us prove that this experiment is further asymptotically equivalent to the experiment generated by the observations

$$z_i' = \sqrt{f(s_i)} + \frac{1}{2}\left(\frac{N}{n}\right)^{1/2} \xi_i'. \tag{38}$$

Let us denote for shortness, $\gamma = (N/n)^{1/2}$. Then

$$\begin{aligned} f_{z_i}(x) &= 2x f_{y_i'}(x^2) \\ &= \sqrt{\frac{2}{\pi}} f^{-1/2}(s_i)\gamma^{-1} x \exp\left\{-\frac{1}{2}\left[f^{-1/2}(s_i)\gamma^{-1}(x^2 - f(s_i))\right]^2\right\} \frac{1}{P(\xi \geq -h)} I_{[0,\infty[}(x). \end{aligned}$$

Also,

$$f_{z_i'}(x) = \sqrt{\frac{2}{\pi}}\gamma^{-1} \exp\left\{-\frac{1}{2}\left[2\gamma^{-1}(x - \sqrt{f(s_i)})\right]^2\right\} \frac{1}{P(\xi \geq -h)} I_{[\sqrt{f(s_i)}/2,\infty[}(x).$$

Write

$$f^{-1/4}(s_i)\left(\sqrt{f(s_i)}+2^{-1}\gamma y\right)^{1/2} = 1-\frac{1}{4}f^{-1/2}(s_i)\gamma y + R^{(1)}(\gamma y)$$

and

$$\begin{aligned}
&\exp\left\{\frac{1}{4}\left(2^{-1}f^{-1/2}(s_i)\gamma y^3+2^{-4}f^{-1}(s_i)\gamma^2y^4\right)\right\}\\
&= 1+\frac{1}{4}\left(2^{-1}f^{-1/2}(s_i)\gamma y^3+2^{-4}f^{-1}(s_i)\gamma^2y^4\right)\\
&\quad +R^{(2)}\left(\frac{1}{4}\left(2^{-1}f^{-1/2}(s_i)\gamma y^3+2^{-4}f^{-1}(s_i)\gamma^2y^4\right)\right)
\end{aligned}$$

where $|R^{(1)}(t)|, |R^{(2)}(t)| \le Const \times t^2$, so that

$$\begin{aligned}
&f^{-1/4}(s_i)\left(\sqrt{f(s_i)}+2^{-1}\gamma y\right)^{1/2}\exp\left\{\frac{1}{4}\left(2^{-1}f^{-1/2}(s_i)\gamma y^3+2^{-4}f^{-1}(s_i)\gamma^2y^4\right)\right\}\\
&\qquad = 1-\frac{1}{4}f^{-1/2}(s_i)\gamma y+\frac{1}{8}f^{-1/2}(s_i)\gamma y^3+R'(\gamma,y)
\end{aligned}$$

where R' is such that

$$\int_{-h}^{\infty}\exp\left\{-\frac{1}{2}y^2\right\}\left|R'(\gamma,y)\right|dy\frac{1}{P(\xi\ge -h)} \le Const\times\gamma^2.$$

Now, making the change of variables $y=2\gamma^{-1}(x-\sqrt{f(s_i)})$,

$$\begin{aligned}
&\frac{1}{2}H^2\left(\mathcal{L}(z_i),\mathcal{L}(z_i')\right)\\
&= 1-\int\sqrt{f_{z_i}f_{z_i'}}\\
&= 1-\int_{\sqrt{f(s_i)}/2}^{\infty}\sqrt{\frac{2}{\pi}}f^{-1/4}(s_i)\gamma^{-1}\sqrt{x}\exp\left\{-\frac{1}{4}\left[f^{-1/2}(s_i)\gamma^{-1}(x^2-f(s_i))\right]^2\right\}\\
&\quad\times\exp\left\{-\frac{1}{4}\left[2\gamma^{-1}(x-\sqrt{f(s_i)})\right]^2\right\}dx\frac{1}{P(\xi\ge -h)}\\
&= 1-\int_{-h}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}y^2\right\}f^{-1/4}(s_i)\left(\sqrt{f(s_i)}+2^{-1}\gamma y\right)^{1/2}\\
&\quad\times\exp\left\{-\frac{1}{4}\left(-2^{-1}f^{-1/2}(s_i)\gamma y^3-2^{-4}f^{-1}(s_i)\gamma^2y^4\right)\right\}dy\frac{1}{P(\xi\ge -h)}\\
&= 1-\int_{-h}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}y^2\right\}\\
&\quad\times\left(1-\frac{1}{4}f^{-1/2}(s_i)\gamma y+\frac{1}{8}f^{-1/2}(s_i)\gamma y^3+R'(\gamma,y)\right)dy\frac{1}{P(\xi\ge -h)}
\end{aligned}$$

$$
\begin{aligned}
&\leq -\int_h^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} \left(-\frac{1}{4}f^{-1/2}(s_i)y + \frac{1}{8}f^{-1/2}(s_i)y^3\right) \gamma dy \frac{1}{P(\xi \geq -h)} \\
&\quad + Const \times \gamma^2 \\
&\leq Const \times \gamma^2 \leq Const \times \frac{N}{n}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&H^2\left(\mathcal{L}((z_i)_{i=1,\ldots,N}), \mathcal{L}((z_i^{'})_{i=1,\ldots,N})\right) \\
&\leq 2NH^2\left(\mathcal{L}(z_1), \mathcal{L}(z_1^{'})\right) \leq Const \times \frac{N^2}{n} \leq Const \times n^{(1+\epsilon_1)/s} n^{-1} \longrightarrow 0.
\end{aligned}
$$

Let us finally prove that the experiment generated by the observations (38) is equivalent to the homoscedastic Gaussian experiment $E_{1,n}$, defuned in (4). This holds because

$$
\begin{aligned}
H^2\left(\mathcal{L}((z_i^{'})_{i=1,\ldots,N}), \mathcal{L}((y_i)_{i=1,\ldots,N})\right) &= 2NH^2\left(\mathcal{L}(z_1^{'}), \mathcal{L}(y_1)\right) \\
&\leq 4N\|\mathcal{L}(z_1^{'}) - \mathcal{L}(y_1)\|_{TV} \\
&= 2N\int \left|f_{z_1^{'}} - f_{y_1}\right| \\
&= 2N\int \left|f_{\xi_1^{'}} - f_{\xi_1}\right| \\
&\leq 4N\int_{-\infty}^{-h} f_{\xi_1} \\
&\leq 2N\exp\left\{-\frac{1}{2}h^2\right\} \longrightarrow 0
\end{aligned}
$$

where we used the fact that $f_{\xi_1^{'}} = f_{\xi_1} I_{[-h,\infty[} / \int_{-h}^{\infty} f_{\xi_1}$. Formula (11) is proved.

4 Continuous White Noise

Let us prove Theorem 3 in this section. That is, we will prove that the continuous Gaussian white noise model is asymptotically equivalent to the homoscedastic discrete Gaussian regression model.

4.1 From Continuous to Discrete

Let us start with proving that the discrete observations in (3) contain asymptotically less information than the continuous white noise model (12). This is the easier direction. Let $K(t) = I_{[-1/2,1/2]}(t)$ and

$$\psi_i(t) = NK(N(t - s_i)) = NI_{[t_{i-1},t_i]}(t)$$

where, as before, $t_i = i/N$ and $s_i = (t_{i-1} + t_i)/2$. Let

$$S_n = \left(\int \psi_i dX_n\right)_{i=1,\dots,N}.$$

Now $Var((Y_n)_i) = Var(y_i) = N/(4n)$ and $Var((S_n)_i) = Var(\int \psi_i dX_n) = \int \psi_i^2/(4n) = N/(4n)$. Thus,

$$\begin{aligned}
\|\mathcal{L}(S_n) - \mathcal{L}(Y_n)\|_{TV}^2 &\leq H^2\left(\mathcal{L}(S_n), \mathcal{L}(Y_n)\right) \\
&\leq 2\sum_{i=1}^{N} H^2\left(\mathcal{L}((S_n)_i), \mathcal{L}((Y_n)_i)\right) \\
&\leq 2\sum_{i=1}^{N}\left[\frac{(\int \psi_i\sqrt{f} - \sqrt{f(s_i)})^2}{n^{-1}N}\right] \\
&\leq 2n \sup_{i=1,\dots,N}\left(\int \psi_i\sqrt{f} - \sqrt{f(s_i)}\right)^2. \qquad (39)
\end{aligned}$$

Let us denote $g = \sqrt{f}$. Then g satisfies the condition (1) with different constant replacing M. Thus we can write

$$g(s_i + t) = g(s_i) + g'(s_i)t + R(t) \qquad (40)$$

where $|R(t)| \leq Const \times |t|^s$. Exactly the same inference as in (35) will lead to

$$\left|\int \psi_i g - g(s_i)\right| \leq Const \times n^{-(1+\epsilon_1)/2}. \qquad (41)$$

Thus, combining (39) and (41),

$$\sup_{f\in\Sigma} \|\mathcal{L}(S_n) - \mathcal{L}(\tilde{Y}_n)\|_{TV} \longrightarrow 0.$$

We have proved that $\delta(E_{n,2}, E_{n,1}) \to 0$.

4.2 From Discrete to Continuous

Let us then prove that the continuous observation (12) contains asymptotically less information than the discrete observations (3). Let us divide the interval $[0,1]$ to $\tilde{n}$ intervals $[v_{j-1}, v_j]$ where

$$v_j = \left[\frac{N}{n^{(1+\epsilon_2)/(2s)}}\right]\frac{j}{N}, \;\; j = 0,\dots,\tilde{n}-1, \;\; v_{\tilde{n}} = 1$$

where $0 < \epsilon_2 < \epsilon_1$ and ϵ_1 was involved in the definition of N. Now the length of one interval is $\kappa^{-1} = \kappa_j^{-1}$,

$$\kappa_j^{-1} = \left[\frac{N}{n^{(1+\epsilon_2)/(2s)}}\right]\frac{1}{N}, \;\; j = 1, \ldots, \tilde{n}-1, \;\; \kappa_{\tilde{n}}^{-1} = 1 - \left[\frac{N}{n^{(1+\epsilon_2)/(2s)}}\right]\frac{\tilde{n}-1}{N}.$$

Also $\tilde{n} = [\kappa] + 1$ and there are

$$[N/n^{(1+\epsilon_2)/(2s)}] \sim N/\kappa \sim n^{(\epsilon-\epsilon_2)/(2s)} \longrightarrow \infty$$

points s_i in the interval $[v_{j-1}, v_j]$, $j = 1, \ldots, \tilde{n}-1$. In the interval $[v_{\tilde{n}-1}, v_{\tilde{n}}] = [v_{\tilde{n}-1}, 1]$ there are $N - (\tilde{n}-1)[N/n^{(1+\epsilon_2)/(2s)}]$ points. Let

$$u_j = (v_{j-1} + v_j)/2, \;\; j = 1, \ldots, \tilde{n}$$

Let

$$K(t) = I_{[-1/2,1/2]}(t), \;\; L(t) = 12tI_{[-1/2,1/2]}(t),$$
$$h_j(t) = \kappa K(\kappa(t-u_j)) = \kappa I_{[v_{j-1},v_j]}(t),$$

and

$$g_j(t) = \kappa^2 L(\kappa(t-u_j)) = \kappa^3 12(t-u_j) I_{[v_{j-1},v_j]}(t),$$

$j = 1, \ldots, \tilde{n}$. Now it holds that $\int K = 1$, $\int tL(t)dt = 1$, $\sum_{i=1}^N (s_i - u_j) h_j(s_i) = 0$, $\sum_{i=1}^N g_j(s_i) = 0$, $\sum_{i=1}^N h_j(s_i) g_j(s_i) = 0$.

Let us move from the discrete observations towards the continuous observation through the following steps. The first step is to transform the observations from (3) to

$$T_n^{(1)} = \left(x_j^{(1)}, z_j^{(1)}\right)_{j=1,\ldots,\tilde{n}} \tag{42}$$

where

$$\begin{aligned} x_j^{(1)} &= \frac{1}{c_1 N}\sum_{i=1}^N h_j(s_i) y_i \\ &= \frac{1}{c_1 N}\sum_{i=1}^N h_j(s_i)\sqrt{f(s_i)} + \frac{1}{2}\left(\frac{N}{n}\right)^{1/2} \frac{1}{c_1 N}\sum_{i=1}^N h_j(s_i)\xi_i, \;\; j = 1, \ldots, \tilde{n}, \end{aligned}$$

$$\begin{aligned} z_j^{(1)} &= \frac{1}{c_2 N}\sum_{i=1}^N g_j(s_i) y_i \\ &= \frac{1}{c_2 N}\sum_{i=1}^N g_j(s_i)\sqrt{f(s_i)} + \frac{1}{2}\left(\frac{N}{n}\right)^{1/2} \frac{1}{c_2 N}\sum_{i=1}^N g_j(s_i)\xi_i, \;\; j = 1, \ldots, \tilde{n}, \end{aligned}$$

where

$$
\begin{aligned}
c_1 &= \frac{1}{N}\sum_{i=1}^{N} h_j(s_i) \\
&= \frac{1}{N}\sum_{\{i: s_i\in[v_{j-1},v_j]\}} \kappa K(\kappa(s_i-u_j)) \\
&= \frac{\kappa}{N}\sum_{\{i: r_{ij}\in[-1/2,1/2]\}} K(r_{ij})
\end{aligned}
$$

where $r_{ij} = \kappa(s_i - u_j)$,

$$
\begin{aligned}
c_2 &= \frac{1}{N}\sum_{i=1}^{N} \kappa(s_i-u_j) g_j(s_i) \\
&= \frac{1}{N}\sum_{\{i: s_i\in[v_{j-1},v_j]\}} \kappa(s_i-u_j)\kappa L(\kappa(s_i-u_j)) \\
&= \frac{\kappa}{N}\sum_{\{i: r_{ij}\in[-1/2,1/2]\}} r_{ij} L(r_{ij})
\end{aligned}
$$

Generally, for every function $G : [-1/2, 1/2] \to \mathbf{R}$ satisfying $G(t) = G(r_{ij})+R(t-r_{ij})$ where $|R(t)| \le Const \times |t|^c$, $0 < c \le 1$, and denoting $q_{yi} = \min\{r_{ij} + \kappa/(2N), 1/2\}$, $q_{ai} = \max\{r_{ij} - \kappa/(2N), -1/2\}$,

$$
\begin{aligned}
&\left|\frac{\kappa}{N}\sum_{\{i: r_{ij}\in[-1/2,1/2]\}} G(r_{ij}) - \int_{-1/2}^{1/2} G\right| \\
&\le \sum_{\{i: r_{ij}\in[-1/2,1/2]\}} \left|\frac{\kappa}{N}G(r_{ij}) - \int_{q_{ai}}^{q_{yi}} G\right| \\
&\le \sum_{\{i: r_{ij}\in[-1/2,1/2]\}} \left|\int_{q_{ai}}^{q_{yi}} R(t-r_{ij})\right| + \frac{2\kappa}{N}\|G\|_\infty \\
&\le Const \times \sum_{\{i: r_{ij}\in[-1/2,1/2]\}} \int_{q_{ai}}^{q_{yi}} |t-r_{ij}|^c dt + \frac{2\kappa}{N}\|G\|_\infty \\
&\le Const \times |q_{y1} - q_{a1}|^c + \frac{2\kappa}{N}\|G\|_\infty \\
&\le Const \times \left(\frac{\kappa}{N}\right)^c \longrightarrow 0,
\end{aligned}
\tag{43}
$$

because $[-1/2, 1/2] = \cup_{\{i: r_{ij}\in[-1/2,1/2]\}}[q_{ai}, q_{yi}]$, $|q_{yi} - q_{ai}| \le \kappa/N$, and $\text{card}\{i : r_{ij} \in [-1/2, 1/2]\} \sim N/\kappa$. Because $\int_{-1/2}^{1/2} K = \int_{-1/2}^{1/2} tL(t) = 1$, we have from (43) that $c_1 \sim c_2 \sim 1$.

The second step is to approximate (42) with

$$T_n^{(2)} = \left(x_j^{(2)}, z_j^{(2)}\right)_{j=1,\ldots,\tilde{n}} \tag{44}$$

where

$$y_j^{(2)} = \frac{1}{c_1 N}\sum_{i=1}^{N} h_j(s_i)\sqrt{f(s_i)} + \frac{1}{2}n^{-1/2}\int h_i dW,\ j = 1,\ldots,\tilde{n}$$

and

$$z_j^{(2)} = \frac{1}{c_2 N}\sum_{i=1}^{N} g_j(s_i)\sqrt{f(s_i)} + \frac{1}{2}n^{-1/2}\int g_i dW,\ j = 1,\ldots,\tilde{n}.$$

The third step is to approximate (44) with

$$T_n^{(3)} = \left(x_j^{(3)}, z_j^{(3)}\right)_{j=1,\ldots,\tilde{n}} \tag{45}$$

where

$$x_j^{(3)} = \int h_j dX_n = \int h_j\sqrt{f} + \frac{1}{2}n^{-1/2}\int h_j dW,\ j = 1,\ldots,\tilde{n}$$

and

$$z_j^{(3)} = \int g_j dX_n = \int g_j\sqrt{f} + \frac{1}{2}n^{-1/2}\int g_j dW,\ j = 1,\ldots,\tilde{n}$$

where X_n was defined in (12).

Let us start with the proof of equivalence of (42) and (44). The random vector $T_n^{(1)}$ is normally distributed with the $2\tilde{n} \times 2\tilde{n}$ covariance matrix

$$C_1 = diag\left(C_{1j}\right)_{j=1,\ldots,\tilde{n}}$$

where

$$\begin{aligned} C_{1j} &= \begin{bmatrix} \frac{1}{nc_1^2 N}\sum_{i=1}^N h_j^2(s_i) & \frac{1}{nc_1 c_2 N}\sum_{i=1}^N h_j(s_i)g_j(s_i) \\ \frac{1}{nc_1 c_2 N}\sum_{i=1}^N h_j(s_i)g_j(s_i) & \frac{1}{nc_2^2 N}\sum_{i=1}^N g_j^2(s_i) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{nc_1^2 N}\sum_{i=1}^N h_j^2(s_i) & 0 \\ 0 & \frac{1}{nc_2^2 N}\sum_{i=1}^N g_j^2(s_i) \end{bmatrix}. \end{aligned}$$

The random vector $T_n^{(2)}$ is normally distributed with the covariance matrix

$$C_2 = diag\left(C_{2j}\right)_{j=1,\ldots,\tilde{n}}$$

where

$$
\begin{aligned}
C_{2j} &= \begin{bmatrix} n^{-1}\int h_j^2 & n^{-1}\int h_j g_j \\ n^{-1}\int h_j g_j & n^{-1}\int g_j^2 \end{bmatrix} \\
&= \begin{bmatrix} n^{-1}\int h_j^2 & 0 \\ 0 & n^{-1}\int g_j^2 \end{bmatrix}
\end{aligned}
$$

where $\int h_j^2 = \kappa$, $= \int g_j^2 = 12\kappa^3$, $\int h_j g_j = \kappa \int g_j = \kappa \int tL(t)dt = 0$. Now, because $\tilde{n} \sim \kappa$,

$$
\begin{aligned}
\|C_2^{-1}\|^2 &= \sum_{j=1}^{\tilde{n}} \|C_{2j}^{-1}\|^2 \\
&= \tilde{n}\left(n^{-1}\int h_1^2\right)^{-2} + \tilde{n}\left(n^{-1}\int g_1^2\right)^{-2} \\
&= \tilde{n}n^2\left(\kappa^{-2} + 12^{-2}\kappa^{-6}\right) \leq Const \times \frac{n^2}{\kappa}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
\|C_1^{-1}\|^2 &= \sum_{j=1}^{\tilde{n}} \|C_{1j}^{-1}\|^2 \\
&= \sum_{j=1}^{\tilde{n}} \left[\left(\frac{1}{nNc_1^2}\sum_{i=1}^{N} h_j^2(s_i)\right)^{-2} + \left(\frac{1}{nNc_2^2}\sum_{i=1}^{N} g_j^2(s_i)\right)^{-2}\right] \\
&\leq Const \times \|C_2^{-1}\|^2 \leq Const \times \frac{n^2}{\kappa}.
\end{aligned}
$$

Thus, using (16) and (43),

$$
\begin{aligned}
&\|\mathcal{L}(T_n^{(1)}) - \mathcal{L}(T_n^{(2)})\|_{TV}^2 \\
&\leq H^2\left(\mathcal{L}(T_n^{(1)}), \mathcal{L}(T_n^{(2)})\right) \\
&\leq \frac{1}{16}\|C_1^{-1}\|\|C_2^{-1}\| \sum_{k,l}([C_1]_{k,l} - [C_2]_{k,l})^2 \\
&\leq Const \times \frac{n^2}{\tilde{n}}\left[\sum_{j=1}^{\tilde{n}}\left(\frac{1}{nc_1N}\sum_{i=1}^{N} h_j^2(s_i) - n^{-1}\int h_j^2\right)^2\right. \\
&\quad \left. + \sum_{j=1}^{\tilde{n}}\left(\frac{1}{nc_2N}\sum_{i=1}^{N} g_j^2(s_i) - n^{-1}\int g_j^2\right)^2\right] \\
&= o(1)
\end{aligned}
$$

Let us move on to the proof of the asymptotic equivalence between (44) and (45). We have that

$$Var(x_j^{(2)}) = Var(x_j^{(3)}) = n^{-1}\int h_j^2 = n^{-1}\kappa$$

and

$$Var(z_j^{(2)}) = Var(z_j^{(3)}) = n^{-1}\int g_j^2 = n^{-1}12\kappa^3.$$

Thus, because $\kappa \sim \tilde{n}$,

$$\begin{aligned}
&\|\mathcal{L}(T_n^{(2)}) - \mathcal{L}(T_n^{(3)})\|_{TV}^2 \\
&\leq \; H^2\left(\mathcal{L}(T_n^{(2)}), \mathcal{L}(T_n^{(3)})\right) \\
&= \; 2\sum_{j=1}^{\tilde{n}} H^2\left(\mathcal{L}(x_j^{(2)}, z_j^{(2)}), \mathcal{L}(x_j^{(3)}, z_j^{(3)})\right) \\
&\leq \; 2\sum_{j=1}^{\tilde{n}}\left[\frac{(E(x_j^{(2)}) - E(x_j^{(3)})^2}{n^{-1}\kappa} + \frac{(E(z_j^{(2)}) - E(z_j^{(3)})^2}{n^{-1}\kappa^3}\right] \\
&\leq \; Const \times n \sup_{j=1,\ldots,\tilde{n}}\left[\left(E(x_j^{(2)}) - E(x_j^{(3)})\right)^2 + \kappa^{-2}\left(E(z_j^{(2)}) - E(z_j^{(3)})\right)^2\right]. \quad (46)
\end{aligned}$$

Denoting $g = \sqrt{f}$, using the decomposition (40) and the fact that $\int tK(t)dt = 0$,

$$\begin{aligned}
E(x_j^{(3)}) &= \int_0^1 h_i g \\
&= \kappa\int_0^1 K(\kappa(t-u_j))g(t)dt \\
&= \int_{-1/2}^{1/2} K(t)g(u_j + \kappa^{-1}t)dt \\
&= \int_{-1/2}^{1/2} K(t)\left(g(u_j) + \kappa^{-1}tg'(u_j) + R(\kappa^{-1}t)\right)dt \\
&= g(u_j) + \int_{-1/2}^{1/2} K(t)R(\kappa^{-1}t)dt.
\end{aligned}$$

Similarily, defining $r_{ij} = \kappa(s_i - u_j)$, using condition (1), and using the fact that $\sum_{\{i:r_{ij}\in[-1/2,1/2]\}} r_{ij}K(r_{ij}) = 0$,

$$\begin{aligned}
&E(x_j^{(2)}) \\
&= \frac{1}{c_1 N}\sum_{i=1}^{N} h_j(s_i)f(s_i) \\
&= \frac{1}{c_1 N}\sum_{\{i:s_i\in[v_{j-1},v_j]\}} \kappa K(\kappa(s_i - u_j))f(s_i)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\kappa}{c_1 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} K(r_{ij}) f(u_j + \kappa^{-1} r_{ij}) \\
&= \frac{\kappa}{c_1 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} K(r_{ij}) \left(g(u_j) + \kappa^{-1} r_{ij} g^{'}(u_j) + R(\kappa^{-1} r_{ij}) \right) \\
&= g(u_j) + \frac{\kappa}{c_1 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} K(r_{ij}) R(\kappa^{-1} r_{ij}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\left| E(x_j^{(2)}) - E(x_j^{(3)}) \right| \\
&= \left| \frac{\kappa}{c_1 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} K(r_{ij}) R(\kappa^{-1} r_{ij}) + \int_{-1/2}^{1/2} K(t) R(\kappa^{-1} t) dt \right| \\
&\leq Const \times \frac{\kappa}{c_1 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} K(r_{ij}) |\kappa^{-1} r_{ij}|^s + \int_{-1/2}^{1/2} K(t) |\kappa^{-1} t|^s dt \\
&\leq Const \times \kappa^{-s} \left[\frac{\kappa}{c_1 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} K(r_{ij}) |r_{ij}|^s + \int_{-1/2}^{1/2} K(t) |t|^s dt \right] \\
&\leq Const \times \kappa^{-s} \leq Const \times n^{-(1+\epsilon_2)/2}.
\end{aligned} \tag{47}
$$

Also, using the fact that $\int L(t)dt = 0$,

$$
\begin{aligned}
E(z_j^{(3)}) &= \int_0^1 g_i g \\
&= \kappa^2 \int_0^1 L(\kappa(t - u_j)) g(t) dt \\
&= \kappa \int_{-1/2}^{1/2} L(t) g(u_j + \kappa^{-1} t) dt \\
&= \kappa \int_{-1/2}^{1/2} L(t) \left(g(u_j) + \kappa^{-1} t g^{'}(u_j) + R(\kappa^{-1} t) \right) dt \\
&= g^{'}(u_j) + \kappa \int_{-1/2}^{1/2} L(t) R(\kappa^{-1} t) dt.
\end{aligned}
$$

Similarily, using the fact that $\sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) = 0$,

$$
\begin{aligned}
&E(z_j^{(2)}) \\
&= \frac{1}{c_2 N} \sum_{i=1}^{N} g_j(s_i) g(s_i) \\
&= \frac{1}{c_2 N} \sum_{\{i: s_i \in [v_{j-1}, v_j]\}} \kappa^2 L(\kappa(s_i - u_j)) g(s_i)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\kappa^2}{c_2 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) g(u_j + \kappa^{-1} r_{ij}) \\
&= \frac{\kappa^2}{c_2 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) \left(g(u_j) + \kappa^{-1} r_{ij} g^{'}(u_j) + R(\kappa^{-1} r_{ij}) \right) \\
&= g^{'}(u_j) + \frac{\kappa^2}{c_2 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) R(\kappa^{-1} r_{ij}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\left| E(z_j^{(2)}) - E(z_j^{(3)}) \right| \\
&= \left| \frac{1}{c_2 N} \sum_{i=1}^{N} g_j(s_i) g(s_i) - \int_0^1 g_i g \right| \\
&\leq \left| \frac{\kappa^2}{c_2 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) R(\kappa^{-1} r_{ij}) + \kappa \int_{-1/2}^{1/2} L(t) R(\kappa^{-1} t) dt \right| \\
&\leq Const \times \frac{\kappa^2}{c_2 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) |\kappa^{-1} r_{ij}|^s + \kappa \int_{-1/2}^{1/2} L(t) |\kappa^{-1} t|^s dt \\
&\leq Const \times \kappa^{1-s} \left[\frac{\kappa}{c_2 N} \sum_{\{i: r_{ij} \in [-1/2, 1/2]\}} L(r_{ij}) |r_{ij}|^s + \int_{-1/2}^{1/2} L(t) |t|^s dt \right] \\
&\leq Const \times \kappa^{1-s} = Const \times \kappa n^{-(1+\epsilon_2)/2}. \qquad (48)
\end{aligned}
$$

From (46), (47), and (48) it follows that

$$
\| \mathcal{L}(T_n^{(2)}) - \mathcal{L}(T_n^{(3)}) \|_{TV} \longrightarrow 0.
$$

Let us finally prove that the continuous white noise model (12) contains asymptotically less information than the observations $T_n^{(3)}$ defined in (45), or actually, that these observations contain asymptotically the same amount information. Let us denote, as before, $g = \sqrt{f}$, and let us denote,

$$
\bar{g}(t) = \sum_{j=1}^{\tilde{n}} \left[g(u_j) + g^{'}(u_j)(t - u_j) \right] I_{[v_{j-1}, v_j[}(t) + 0 I_{\{1\}}(t).
$$

Let

$$
\tilde{E}_{n,2} = \left(C[0,1], \mathcal{B}_{C[0,1]}, (Q_{\bar{g},n}^{(2)}, f \in \Sigma) \right)
$$

where $Q_{g,n}^{(2)}$ is the distribution of the continuous white noise observation (12). Now

$$
\frac{dQ_{\bar{g},n}^{(2)}}{dQ_{0,n}^{(2)}}(y) = \exp \left\{ \int \bar{g} dy - \frac{1}{2} \int \bar{g}^2 \right\}
$$

and

$$\int \bar{g}dy = \sum_{j=1}^{\tilde{n}} \left[g(u_j)(y(v_j) - y(v_{j-1})) + g'(u_j) \int_{v_{j-1}}^{v_j} (t - u_j) dy(t) \right].$$

Thus $T_n^{(3)}$ is sufficient statistics in $\tilde{E}_{n,2}$. Thus the experiment generated by the observations $T_n^{(3)}$, defined in (45), is equivalent to the experiment $\tilde{E}_{n,2}$. Finally, by (18),

$$\begin{aligned}
H^2\left(Q_{g,n}^{(2)}, Q_{\bar{g},n}^{(2)}\right) &\leq Const \times n \int (g - \bar{g})^2 \\
&= Const \times n \sum_{j=1}^{\tilde{n}} \int_{v_{j-1}}^{v_j} \left[g(t) - g(u_j) - g'(u_j)(t - u_j) \right]^2 dt \\
&\leq Const \times n \sum_{j=1}^{\tilde{n}} \int_{v_{j-1}}^{v_j} (v_j - v_{j-1})^{2s} dt \\
&\leq Const \times n\tilde{n}^{-2s} \leq Const \times n^{-\epsilon_2} \longrightarrow 0.
\end{aligned}$$

Thus

$$\Delta\left(\tilde{E}_{n,2}, E_{n,2}\right) \longrightarrow 0 \tag{49}$$

We have proved that $\delta(E_{n,1}, E_{n,2}) \to 0$.

References

[1] Bretagnolle, J. and Massart, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Statist.* **17** 239–256.

[2] Brown, L. D. and Low, M. G. (1996a). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

[3] Brown, L. D. and Low, M. G. (1996b). Personal communication, Stochastic Dynamics and Nonparametric Statistics, Berlin, 1996.

[4] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion) *J. Roy. Statist. Soc. B* **57** 301–369.

[5] Golubev, G. K. and Nussbaum, M. (1998). Asymptotic equivalence of spectral density and regression estimation. Manuscript.

[6] Grama, I. and Nussbaum, M. (1996). Asymptotic equivalence of nonparametric generalized linear models. Preprint No. 289, Weierstrass Institute.

[7] Holmström, L. and Klemelä, J. (1992). Asymptotic bounds for the expected L^1 error of a multivariate kernel density estimator. *J. Multivariate Anal.* **42** 245–266.

[8] Laredo, C. F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *Ann. Statist.* **18** 1158–1171.

[9] Le Cam, L. (1969). *Théorie Asymptotique de la Décision Statistique.* Univ. of Montreal Press.

[10] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

[11] Müller, D. W. (1981). The increase of risk due to inaccurate models. *Symposia Mathematica.* Institute Nazionale di alta Mathematica, Vol. 25.

[12] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

[13] Shiryaev, A. N. and Spokoiny, V. (1993). *Statistical Experiments and Decisions: Asymptotic Theory.* To appear.

[14] Strasser, H. (1985). *Mathematical Theory of Statistics.* de Gruyter, Amsterdam.

[15] Talagrand, M. (1988). Small tails for the supremum of a Gaussian process. *Ann. Inst. Henri Poincaré* **24** 307–315.